Peter M. Bluhm
566 East Hill Road
Middlesex, Vermont 05602

October 18, 2011

VIA ELECTRONIC FILING
Ms. Marlene H. Dortch, Secretary
Federal Communications Commission
445 12th Street SW
Washington, DC 20554

Re: National Broadband Plan for Our Future, GN Docket No. 09-51, Connect America Fund, WC Docket No. 10-90, High-Cost Universal Service Support, WC Docket No. 05-337, and Federal-State Joint Board on Universal Service, CC Docket No. 96-45; Lifeline and Link-Up, WC Docket No. 03-109; Universal Service Contribution Methodology, WC Docket No. 06-122.

Dear Ms. Dortch:

I am writing to suggest a simplification to the Windstream regression cost model, as well as point out some concerns with the model. I understand that the Commission may use this model on an interim basis during 2012 while it continues on the CQBAT cost model. I base my comments on a letter in the record filed by Jennie B. Chandra of Windstream dated June 30, 2011.

Windstream prepared its regression-based model by analyzing the detailed cost outputs from the Commission's own HCPM or "synthesis" cost model adopted in the late 1990s. I note initially that the Windstream model is a "model upon a model" in that it is based upon an older forward-looking cost model used to distribute universal service support. If one assumes that a forward-looking model should be used for universal service (as opposed to actual investment), I see nothing wrong in principle with using a regression-based model as opposed to an engineering-type model.

The Windstream regression equation is loglinear, taking the following form:

$$\ln TC = 7.08 + 0.02 * \ln DistCO - 0.15 * \ln Locs + 0.22 * \ln R + 0.06 * \ln Locs^2 - 0.01 * \ln BL^2 - 0.07 * \ln \frac{Locs}{A}$$

Where:

- Total cost = TC
- Distance to central office (in feet) = DistCO
- Total locations = (residential and business locations)=Locs
- Roadfeet in CLLI =R
- Business locations = BL
- LocDensity = locations / area = Locs/A

My first concern is that this model is needlessly complex, making it difficult to understand how the input variables affect the outcome.

- Two terms (Locs and BL) are squared before they are turned into logarithms.
- The input data are not independent. Locations data is used by itself and as a component of the density term.
- Locations appears three times: once with a negative coefficient; a second time squared with a positive coefficient; and a third time inside the density term with a negative coefficient.

The equation can be simplified algebraically as follows:

$$\ln(TC) = 7.08 + 0.02 \ln DistCO + 0.22 \ln R - 0.10 ln\ Locs - 0.02 \ln(BL) + 0.07 \ln A$$

The derivation of this simplified equation is attached on a separate page. This simpler equation produces the same mathematical result as the Windstream equation, but it has several advantages:

- Each independently measured input variable appears once. The exception is that business locations still appears in both the Locs and BL terms.
- The squared terms are gone. As a matter of algebra, $\ln(x^2) = 2*\ln(x)$, so the squared terms are an unnecessary complication.
- The three terms that use locations have been consolidated into one. As a consequence, the area portion of Windstream's density term survives simply as area.

By eliminating multiple location terms appearing with coefficients of different sign and with different exponents, the revised equation is simpler to interpret and evaluate. The result can now be analyzed one variable at a time and thereby evaluated more easily for plausibility. As in the Windstream filing, all costs are expressed on a per month basis. The analysis follows:

1. The first term, the fixed 7.08, produces a cost even for a wire center with minimal locations, distribution routes, and other cost drivers. Where all the inputs are set at 1.0, the minimum configuration cost appears to be $1,188 per month. A minimum configuration central office does have real estate and switching costs, so a nonzero figure seems plausible.
2. The second term is CO Distance. Changing the set of minimum input values to make CO Distance equal to 52,800 (ten miles) adds $289 to the minimum configuration monthly cost, an increase of 24%. This result appears plausible.
3. The third term measures road feet in the exchange area. Changing the set of minimum input values to make Road Feet equal to 52,800 (ten miles) adds $11,807 of monthly cost, increasing the minimum configuration cost by 994%. This result also appears plausible because, as Windstream notes, most of the cost of a telephone exchange is in outside plant.
4. The fourth term is locations. This term's negative coefficient means that an increase in locations reduces cost. Changing the set of minimum input values to make Locations equal to 1,000 reduces the minimum configuration cost by 50%. It is difficult to understand how the incremental cost of an additional location could be negative.
5. The fifth term is business locations. It also has a negative coefficient. Changing the set of minimum input values to make Business Locations equal to 1,000 reduces the minimum configuration cost by 56%. This effect appears difficult to understand for the same reasons as in the preceding paragraph.
6. The sixth term is area. Changing the set of minimum input values to make area equal to 100 (ten miles on a side) adds $452 of monthly cost, increasing the minimum configuration cost by 38%. This result appears plausible, although it is difficult to evaluate the cost effects of area alone, independent of road mileage and locations to be served.

The analysis supporting these observations is attached on a separate page. I understand that it is not always possible to assess the dynamics of a multivariate regression on a term-by-term basis. Seeing a coefficient take an unexpected sign is not necessarily an indication that there is an error. It does, however, require some explanation, and I have did not see any such explanation in the Windstream filing.

I offer two more general observations regarding the use of the log-linear method of analysis. First, a log-linear regression may not be accurate for low input values. For example, if an exchange has zero business locations, the regression calculation cannot be performed because the logarithm of zero is undefined. For at least some of the variables (like road length in the exchange area), the value of the variable will probably never be small. But other variables, such as business locations, could well take small values, and the log-linear regression could perform unpredictably. For example, starting with the minimum configuration, adding ten business locations decreases cost by $56. Adding a second set of ten locations only decreases cost by $14. Thus, at small values, the equation may behave much differently than at large values. A better way to handle this problem may be to create a series of linear equations with breakpoints, much as the Commission has been considering for limiting corporate operations expense.

A second concern about log-linear regression is that in the resulting cost equation, the variables do not act independently, but create a synergistic cost effect. In general, if:

$$ln(Cost) = a * ln(x1) + b * ln(x2)$$

then:

$$Cost = x1^a * x2^b$$

The terms in this type of equation are multiplied, not added. Therefore, the size of any term tends to affect the cost generated by other terms derived from independent data. For example, increasing the CO Distance to ten miles increases the minimum configuration cost by $289. Increasing the road feet to ten miles increases the minimum configuration cost by $11,807. Increasing area to 100 square miles increases the minimum configuration cost by $452. The sum of these three separate increases is $12,548. Yet increasing all three factors at the same time would increase the minimum configuration cost by $21,109. The synergistic effect is $8,561, a substantial proportion of the total.

One should not conclude from this that $21,109 is an inaccurate or inflated cost estimate. The problem, if any, arises from trying to understand and explain. It seems difficult to understand how such powerful cost changes could actually arise in a real network due to synergies among these independent variables. Log-linear regressions are a standard statistical technique, and this synergy behavior may be unavoidable in any analysis using this technique,. Moreover, since the revised equation has five independent variables, synergistic interactions can be sizeable, complex and difficult to predict.

In sum:

- The Windstream regression equation is needlessly complex, and can be simplified to facilitate analysis and evaluation.
- The negative coefficients in the Windstream equation for locations and business locations are difficult to understand and should be carefully evaluated.
- The use of log-linear regression creates some risks of inaccuracy from low values and from synergistic interactions among the terms.

These problems do not necessarily indicate that the Windstream regression was inaccurate. Rather, they suggest areas for further study. Ideally, the Commission would allow the public access to the underlying data, so that others can evaluate the regression methodology that Windstream has suggested, and see if improvements can be made. If that public input is not possible at this point, then I hope these comments can provide a useful direction for the Commission's own investigations.

Sincerely,

Peter Bluhm
566 East Hill Road
Middlesex, Vermont 05602
802 229-0130

Cc: Sharon Gillett
Carol Mattey
Katie King
Steve Rosenberg

Attachment 1: Derivation of Simpler Regression Formula

Derivation of Restated Windstream Regression Formula:

The variables in the Windstream regression equation are:
- Distance to central office (in feet) = DistCO
- Total locations = (residential and business locations)=Locs
- Roadfeet in CLLI =R
- Business locations = BL
- LocDensity = Total locations/Area = Locs/A

The coefficients from the Windstream regression study are:
- a=7.08
- b=0.02
- c=-0.15
- d=0.22
- e=0.06
- f=-0.01
- g=-0.07

The loglinear regression model from Windstream takes the following form:

$$\ln(TC) = 7.08 + 0.02 * \ln\ DistCO\ - 0.15 * \ln(Locs) + 0.22 * \ln\ R\ + 0.06 * \ln\ Locs^2 \\ - 0.01 * \ln\ BL^2\ - 0.07 * \ln\ \frac{Locs}{A}$$

Using the properties of logarithms:

$$\ln(TC) = 7.08 + \ln\ DistCO^{0.02} * Locs^{-0.15} * R^{0.22} * Locs^{0.12} * BL^{-0.02} * \ \frac{Locs}{A}^{-0.07}$$

Clarifying the last term:

$$\ln(TC) = 7.08 + \ln\ DistCO^{0.02} * Locs^{-0.15} * R^{0.22} * Locs^{0.12} * BL^{-0.02} * Locs^{-0.07} * A^{0.07}$$

It is now possible to simplify, consolidating the three Locs terms by adding the exponents:

$$\ln(TC) = 7.08 + \ln\ DistCO^{0.02} * Locs^{-0.10} * R^{0.22} * BL^{-0.02} * A^{0.07}$$

From the properties of logarithms:

$$\ln(TC) = 7.08 + 0.02 \ln\ DistCO\ - 0.10 ln\ Locs\ + 0.22 \ln\ R\ - 0.02 \ln(BL) + 0.07 \ln\ A$$

Rearranging:

$$\ln(TC) = 7.08 + 0.02 \ln\ DistCO\ + 0.22 \ln\ R\ - 0.10 ln\ Locs\ - 0.02 \ln(BL) + 0.07 \ln\ A$$

Attachment 2: Input Variations Using Simpler Formula

	Input Values				
Case #	CO Dist	Roadfeet	Locations	Business Locations	Area
1	1	1	1	1	1
2	52800	1	1	1	1
3	1	52800	1	1	1
4	1	1	1000	1	1
5	1	1	1000	1000	1
6	1	1	1	1	100
7	52800	52800	1	1	100
8	1	1	1	11	1
9	1	1	1	21	1

	Cost Effects									
Term #	1	2	3	4	5	6				
Coefficient	7.08	0.02	0.22	(0.10)	(0.02)	0.07				
Case #	Intcpt	CO Dist	Rd Ft	Locs	BL	Area	Log Cost	Total Monthly Cost	Cost Change from case #1	Change %
1	7.08	-	-	-	-	-	7.08	$ 1,188		
2	7.08	0.22	-	-	-	-	7.30	$ 1,477	$ 289	24%
3	7.08	-	2.39	-	-	-	9.47	$ 12,995	$ 11,807	994%
4	7.08	-	-	(0.69)	-	-	6.39	$ 595	$ (593)	-50%
5	7.08	-	-	(0.69)	(0.14)	-	6.25	$ 519	$ (669)	-56%
6	7.08	-	-	-	-	0.32	7.40	$ 1,640	$ 452	38%
7	7.08	0.22	2.39	-	-	0.32	10.01	$ 22,297	$ 21,109	1777%
8	7.08	-	-	-	(0.05)	-	7.03	$ 1,132	$ (56)	-5%
9	7.08	-	-	-	(0.06)	-	7.02	$ 1,118	$ (70)	-6%